

Calculating the relative abundances of STs from metagenomic data

JJ Jay-Hyun Jo HK Heidi H. Kong

Updated date: Jan 12, 2022

An abbreviated version of this protocol was published in Science Translational Medicine in Dec 2021

Alterations of human skin microbiome and expansion of antimicrobial resistance after systemic antibiotics

DOI: 10.1126/scitranslmed.abd8077

Detailed protocol

Approximation of relative abundance of each sequence type (ST) from metagenomic data was performed using the workflow called 'Bayesian Identification of Bacteria (BIB)' (1), and followed recommended workflow from the author's github (<https://github.com/PROBIC/BIB>).

First, the genome assembly of each ST and metagenomic data are required. As an initial step, core-alignment of all the STs need to be extracted. This process requires progressiveMauve (2).

```
$ progressiveMauve --output=full_alignment.xmfa ST_X_assembly.fasta ST_Y_assembly.fasta ST_Z_assembly.fasta
$ stripSubsetLCBs full_alignment.xmfa full_alignment.xmfa.bbcolls core_alignment.xmfa 500 4
```

And change xmfa format to fasta, and remove gaps, which are generated during the alignment.

```
$ perl xmfa2fasta.pl --file core_alignment.xmfa > core_alignment.fasta
$ sed 's/-/g' core_alignment.fasta > core_alignment_gapless.fasta
```

Once the fasta-formatted core alignment is ready, the metagenome data need to be aligned to the core alignment (core_alignment_gapless.fasta), using Bowtie2 (3).

```
$ bowtie2-build core_alignment_gapless.fasta core_alignment_gapless
$ bowtie2 -x core_alignment_gapless -U metagenome_reads.fastq -S metagenome_aligned.sam -a
```

Then, estimate the abundances of different STs using the alignment (metagenome_aligned.sam) and the core alignment (core_alignment_gapless.fasta), using BitSeq (4)

```
$ parseAlignment metagenome_aligned.sam -o alignment_info.prob --trSeqFile core_alignment_gapless.fasta --trInfoFile genome_info.tr --uniform --verbose
$ estimateVBExpression -o final_abundance_info.prob -t genome_info.tr
```

After successfully finish the pipeline, user should have a file named 'final_abundance.m_alphas', which illustrates the abundance of each ST in the metagenome data.

References

1. Sankar A, et al. Microb Genom. 2016.
2. Darling AE, Mau B, and Perna NT. PLoS ONE. 2010.
3. Langmead B, and Salzberg SL. Nat. Methods. 2012.

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Jo, J. and Kong, H. (2022). Calculating the relative abundances of STs from metagenomic data. Bio-protocol Preprint. bio-protocol.org/prep1505.
2. Jo, J., Harkins, C. P., Schwardt, N. H., Portillo, J. A., Program, N. C. S., Zimmerman, M. D., Carter, C. L., Hossen, M. A., Peer, C. J., Polley, E. C., Dartois, V., Figg, W. D., Moutsopoulos, N. M., Segre, J. A. and Kong, H. H. (2021). Alterations of human skin microbiome and expansion of antimicrobial resistance after systemic antibiotics. Science Translational Medicine 13(625). DOI: [10.1126/scitranslmed.abd8077](https://doi.org/10.1126/scitranslmed.abd8077)

Copyright: Content may be subjected to copyright.